Capstone Project
Hotel Booking Analysis
By

Kunal Shinkar

Omkar Katare

Ganesh Ghadge

Index

Data Summary

Numeric	Categorical	Binary	Date Time
required_car_parking_spaces	Hotel	Is cancelled	Reservation Status Date
previous_bookings_not_canceled	Country	Is repeated guest	
previous_bookings_not_canceled	Market segment		
stays_in_week_nights	Meal		
stays_in_week_nights	Arrival Date Month		
stays_in_weekend_nights	Distribution Channel		
previous_cancellations	Assigned room type		
Lead time	Reserved room type		
Arrival Date Year	Deposit Type		
Arrival Date Week	Reservation Status		
Arrival date day of month			
adults			
Arrival date week number			
babies			
children			
booking_changes			
total_of_special_requests			
adr			

Null Values Present in Dataset

```
company      94.306893
agent        13.686238
country       0.408744
children      0.003350
dtype: float64
```

1. This Dataset consist four columns which contains null values.

2. For dealing with this null values we drop company and agent column because it contains large amount of Null values.

3. Country and Children contains very small amount of Null values so that we drop rows from these columns which contains null values.

Number of bookings cancelled and its percentages of cancellation

From total number of bookings around 37 % bookings were cancelled.
Confirm bookings was around 63%.

Total bookings = 119206
Confirm bookings = 75011
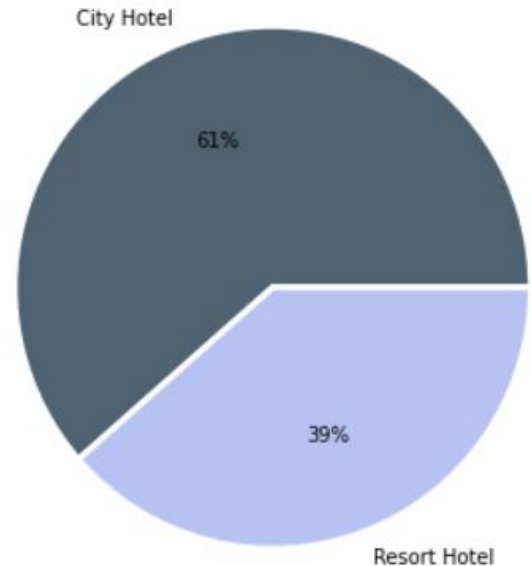Cancelled bookings = 44195
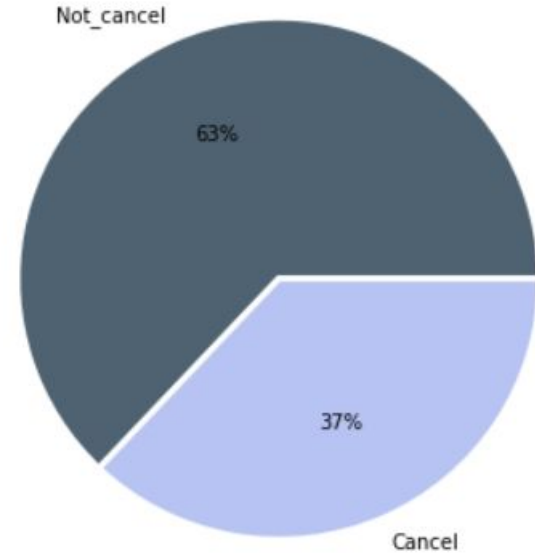
Types of Hotels present in Dataset and most preferred among those by the customers

In this Dataset two types of hotel:
1) City Hotel
2) Resort Hotel
From total bookings around 61% bookings came for City hotel & 39% bookings for Resort Hotel
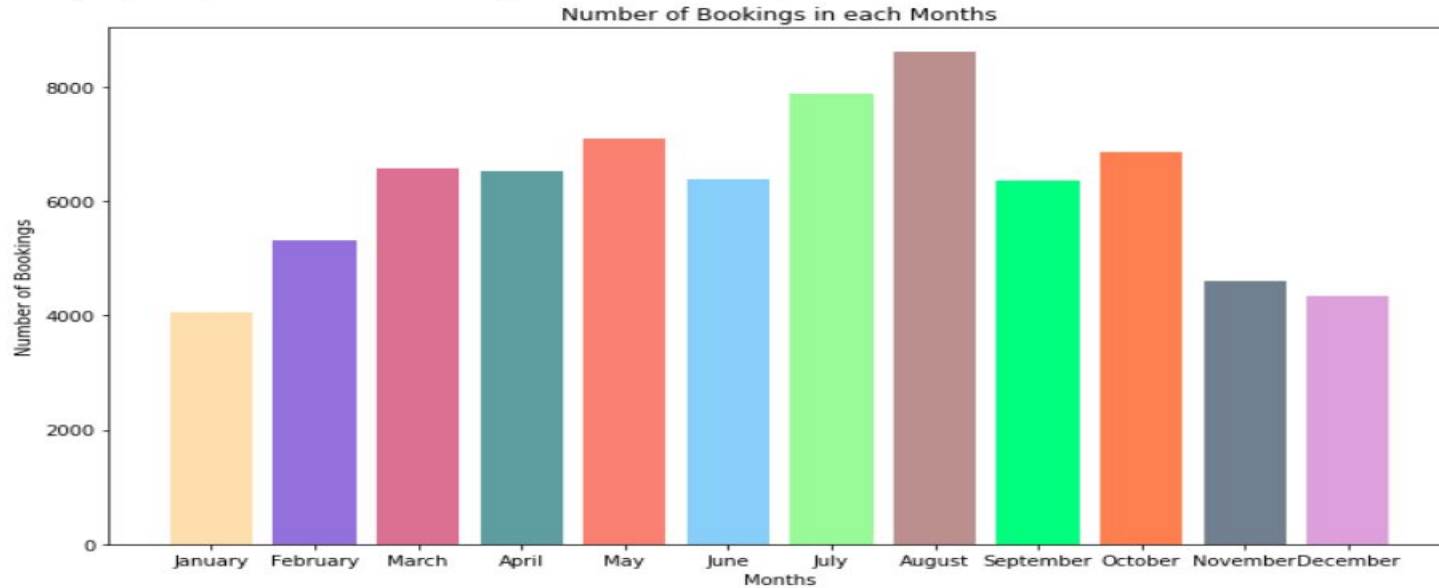
Number of bookings cancelled across both type of hotels and its percentages of cancellation

1. From this pie chart we can see that cancelling rate is 75% in City hotel

2. Cancelling rate in Resort hotel is less as compared to city hotel
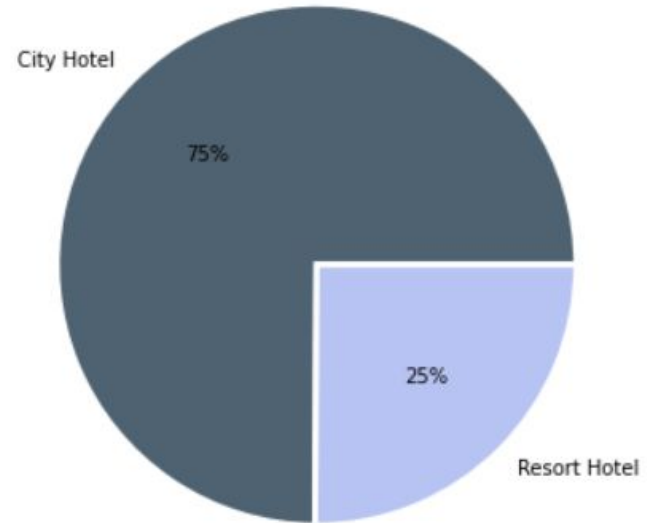
Which is the most trending months of hotel bookings ?


Number of Bookings in each Months

1. Number of Bookings in August and July month is more as compared to other months
2. From month march to October number of bookings was pretty good
3. Number of Bookings decreases in last months of the year

Monthly Bookings by each type of Hotels


Percentage of Bookings in each Months
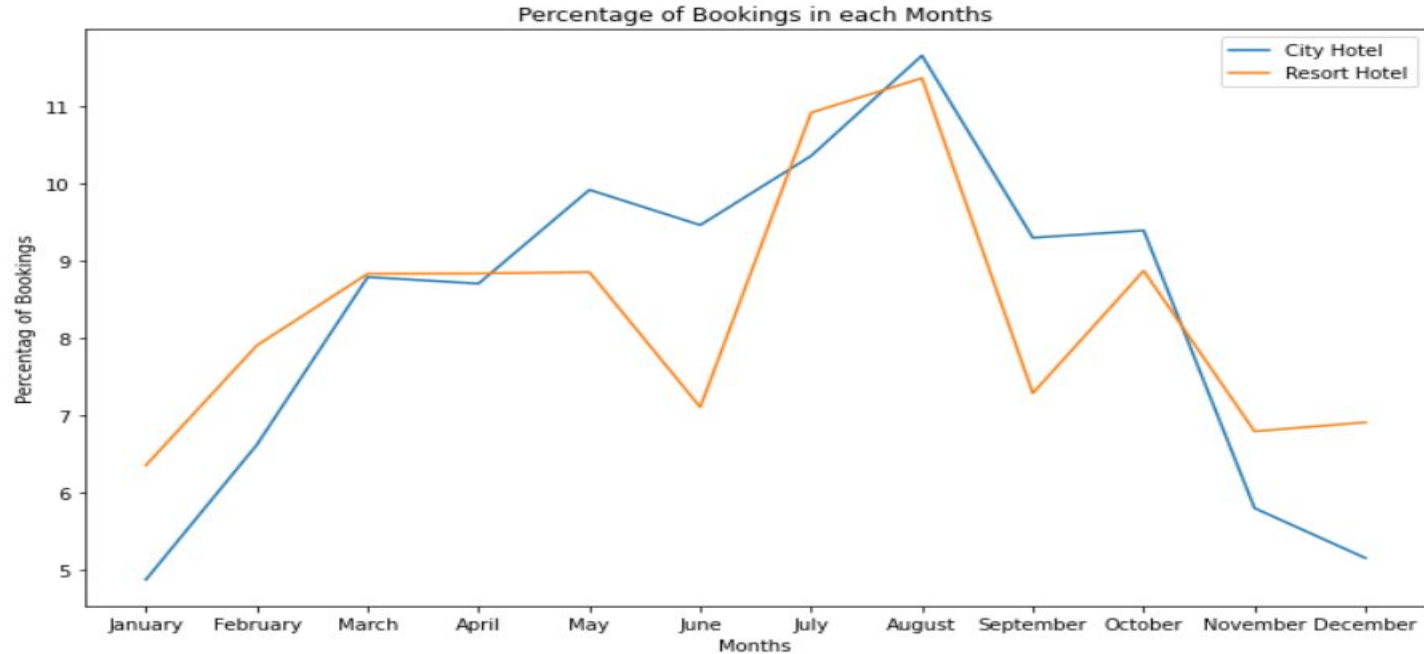
1. Number of bookings for resort hotel constant for March , April and may months.
2. For months may to September bookings for city hotel is more as compared to resort hotel

Number of Bookings across the years
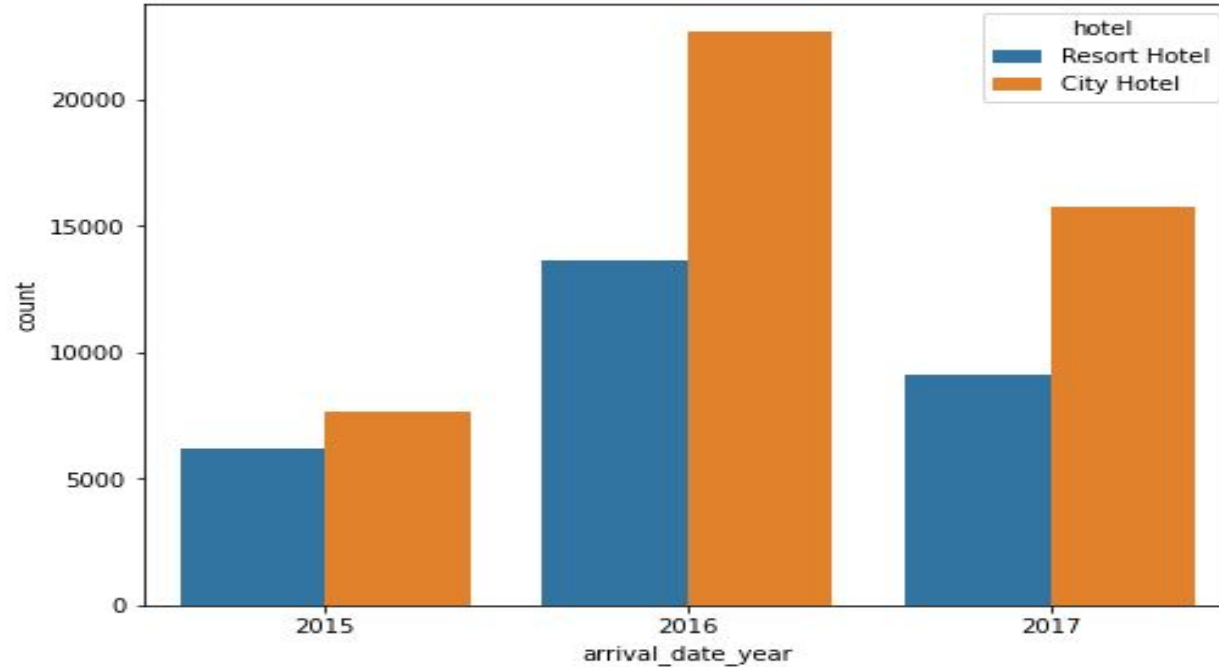
The above Countplot between number of confirmed bookings and year of bookings shows that the number of confirmed bookings was lowest in the year 2015 and was highest in the year 2016. In year 2017 we can see a dip in the number of confirmed bookings relative to year 2016. Throughout these years City Hotels had larger share in the bookings received.

Month wise Number of waiting Days


Average Waiting Time for each month
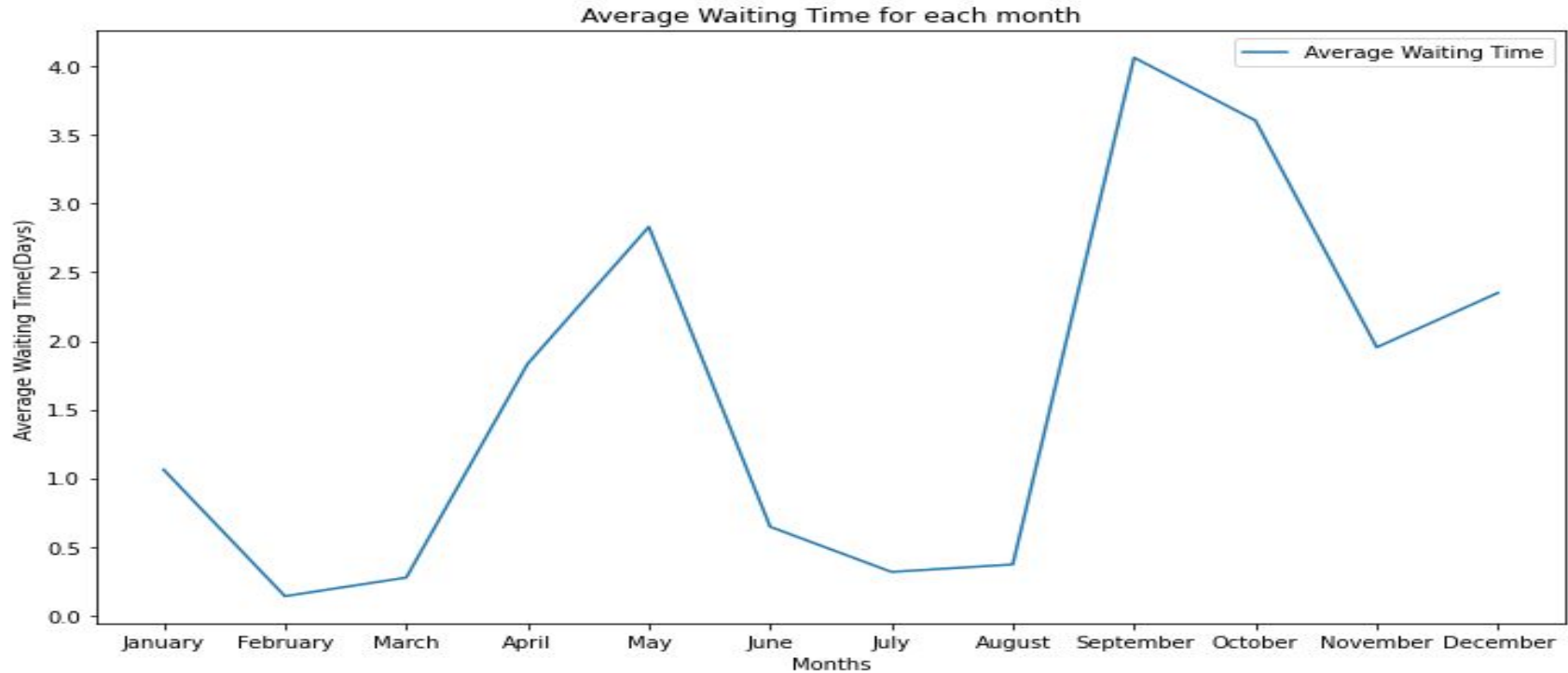
The above line plot shows that guests had to face least waiting time in the months of February, June , July. And the waiting time was highest in the month of August.

Meals ordered by customers
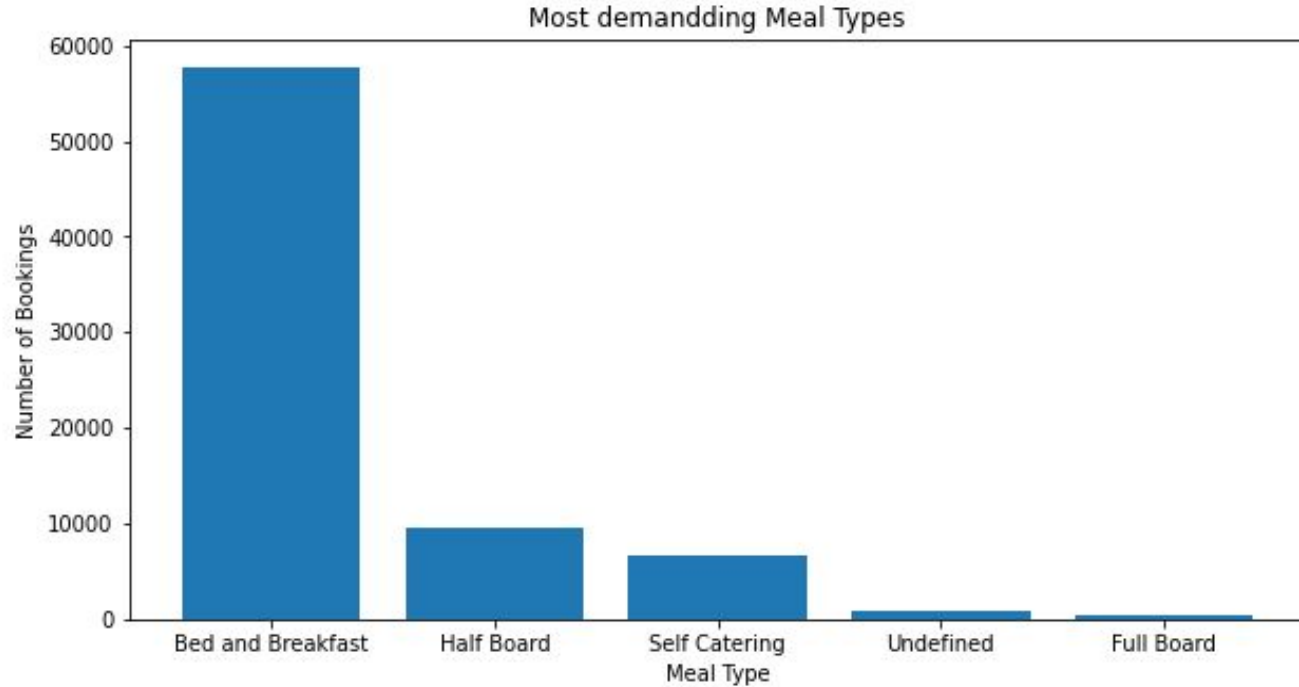
The above Bar Graph plot shows that 'Bread and Breakfast' was the most ordered meal type.

# Most preferred stay duration

Hotel Bookings across the contries
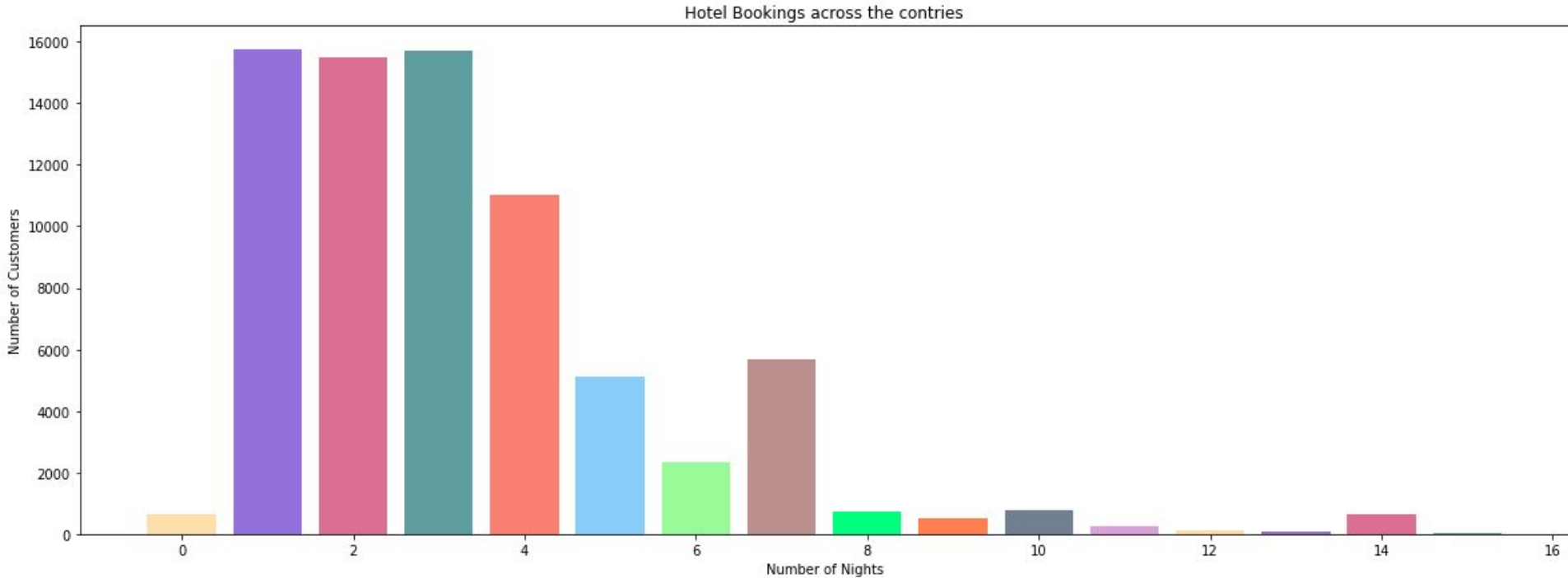
The above bar graph plot shows that most guests prefer to stay in the hotels for 1 ,2, 3 or 4 nights. There are considerable number of bookings for 7 nights stay.

Average daily rates according to room type


Price of rooms according to room types.
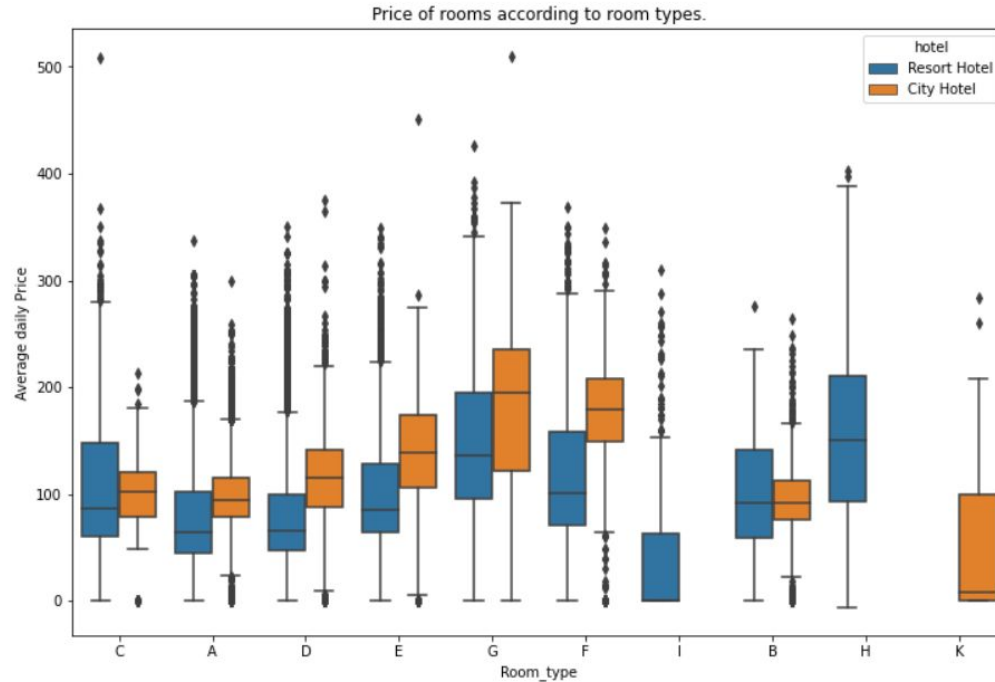
We can see that mean ADR were highest for the room type-G and were lowest for room type-K in City Hotels. In case of Resort Hotels mean ADR was highest for room type-H and was lowest for room type-I

# Month and year wise mean ADR

AVERAGE ADR OF MONTH & YEAR
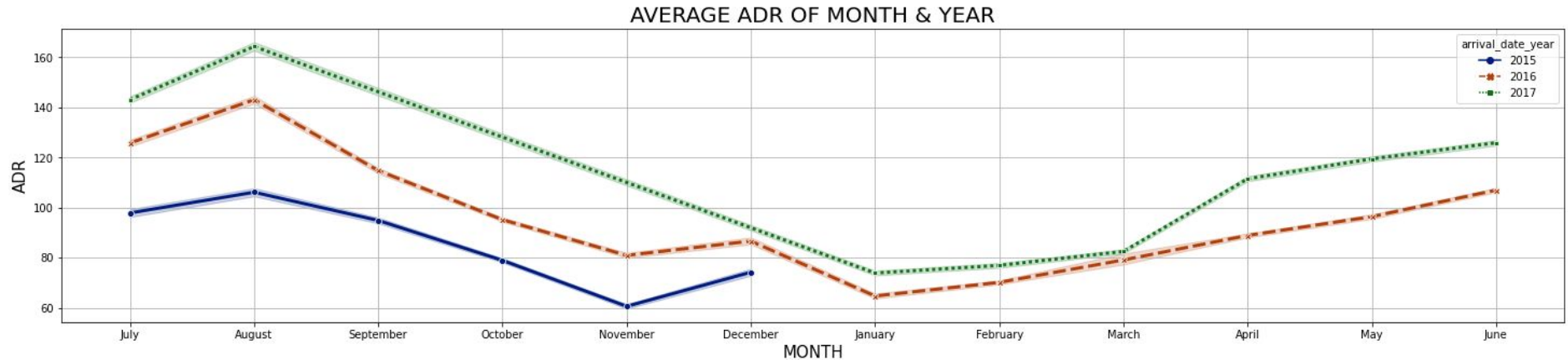
From above plot we can see that ADR was highest in the month of August and it follows a decreasing trend till the month of January from where ADR starts to increase till the month of August . Also on comparing Year on Year basis we can observe that ADR has been consistently increasing.

Lead Time vs Cancellation
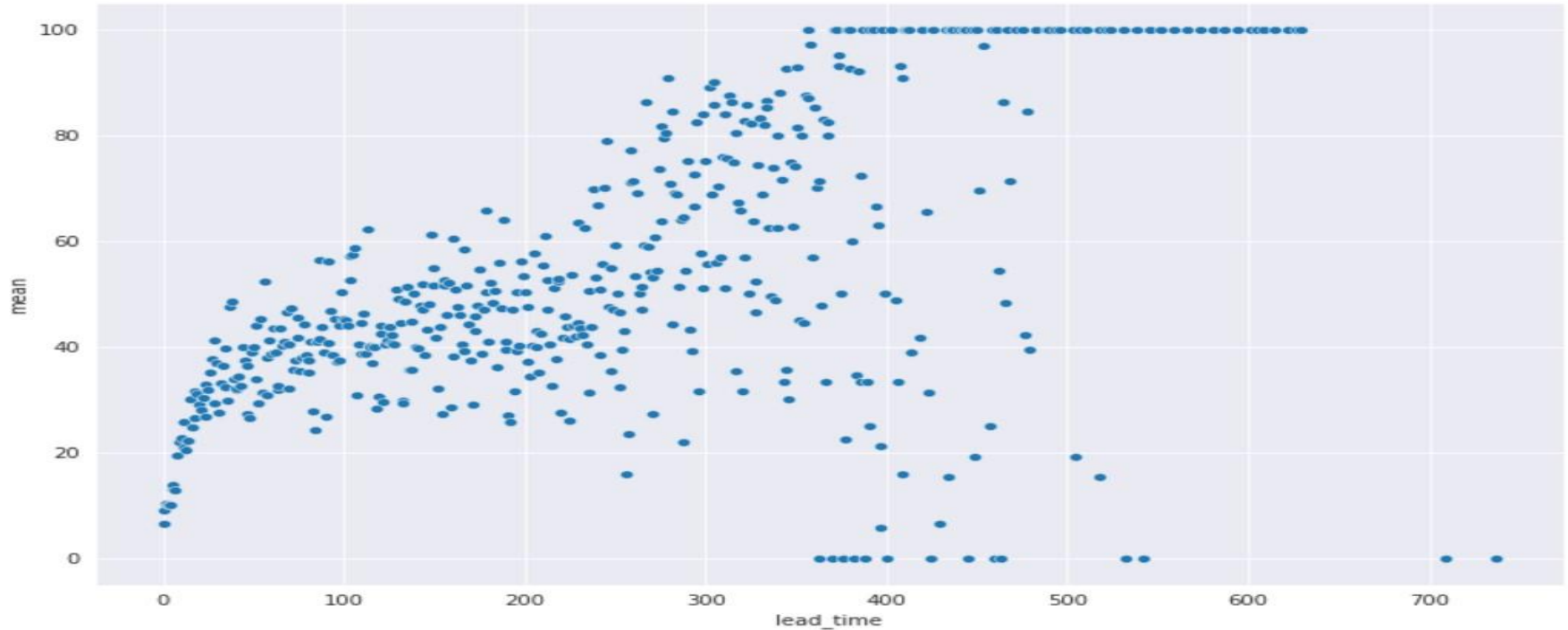
From the above scatter plot, we can see that the higher lead time has higher chance of cancellation. We have plotted lead_time in X direction and the mean of is_cancelled groupby lead_time in Y direction.

Country-wise Hotel Booking

Hotel Bookings across the contries
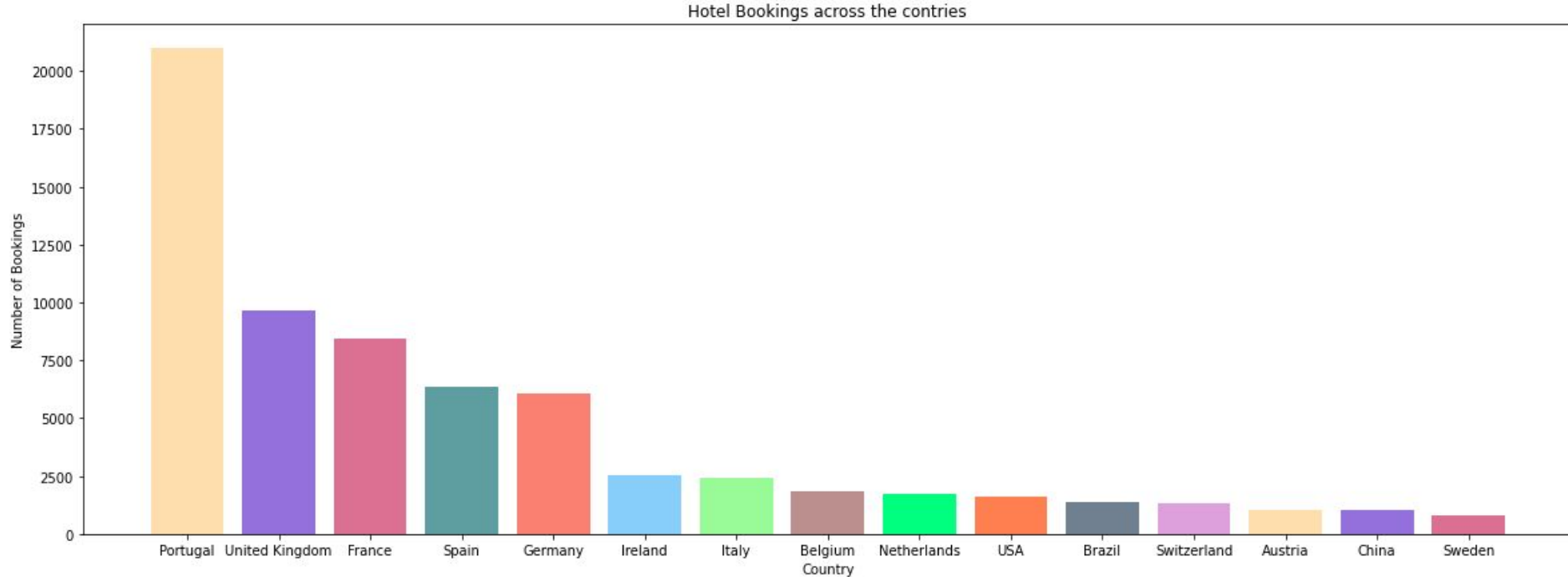
Guests where coming from all over the world throughout the year from 2015 to 2017. The above bar graph shows the top 15 countries from where the guests come. Portugal brings in the mosts numbers of guests.

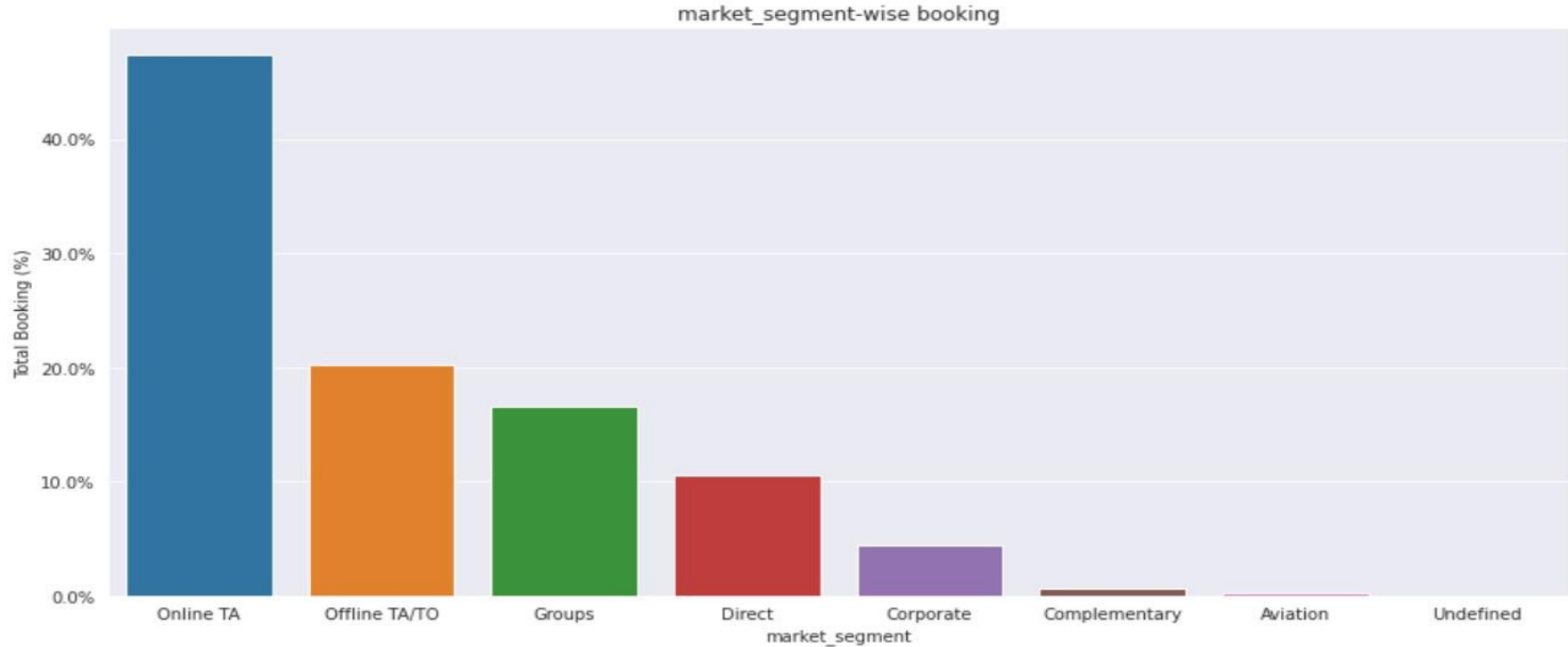The above choropleth map shows that the European region collectively brings in the most number of bookings and the African region collectively brings in the least number of bookings.

Market Segment-wise Hotel Booking Percentage


market_segment-wise booking

Inference: Online Travel Agents brings in more than 40% of the total bookings, followed by the Offline Travel Agents/Tour Operators brings in more than 20% of the total bookings.

Correlation Heat Map Between All Features

AI

Correlation Conclusion:

1. arrival_date_week_number and arrival_date_year are 54% negative correlated.

2. previous_bookings_not_cancelled and is_repeated_guest are 42% positive correlated.

3. Adr and children are positive correlated by 33%.

Correlation Heatmap
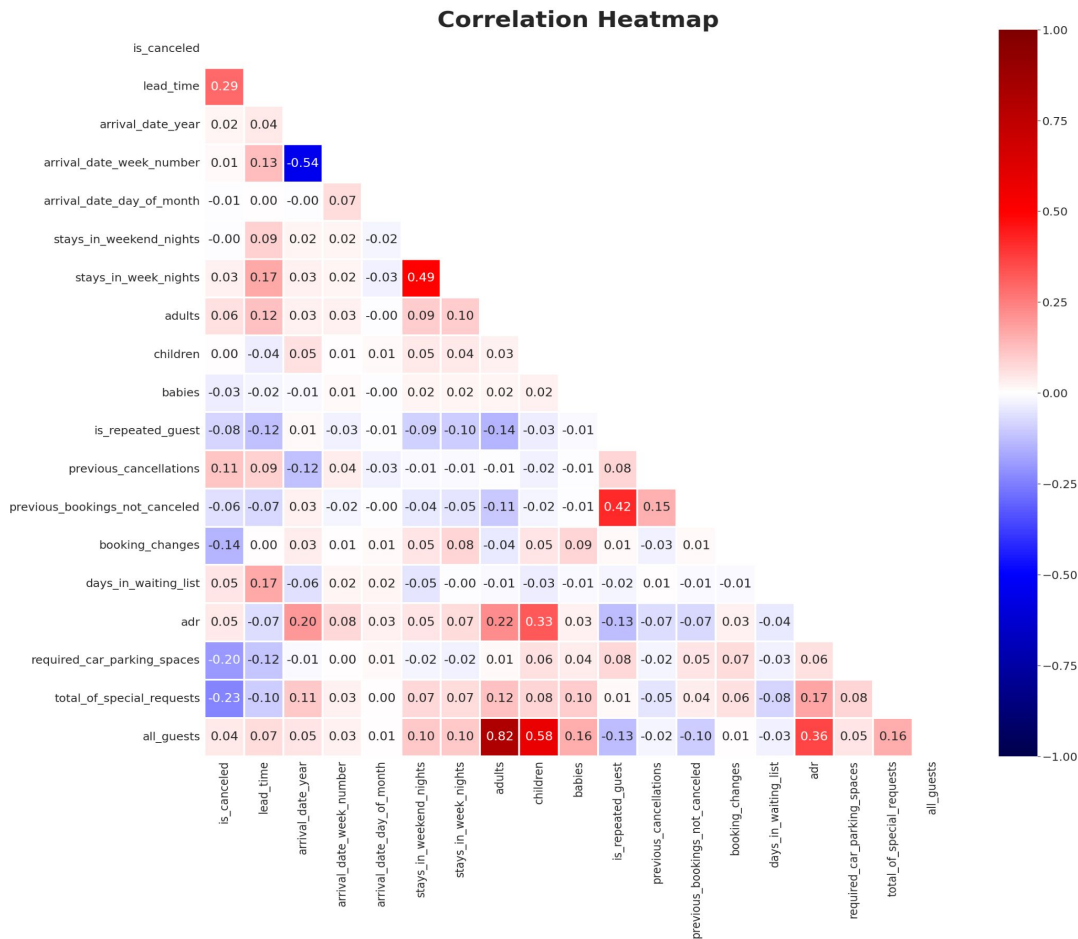
Conclusion:

1. 'City hotels' and 'Resort hotels' are two types of hotels present in the dataset. 'City hotels are more preferred by the customers than the 'Resort Hotel' (66.4% customer prefers 'City hotels' whereas 33.6% customer prefers 'Resort hotels').
2. 'August' month has highest number of hotel bookings whereas 'January' records the least.
3. Guests from PRT(Portugal) have got the maximum number of hotel bookings.
4. Its observed that 'City hotels' were more cancelled as compared to 'Resort hotels'.
5. 'Online TA' brings maximum bookings.
6. 'August' has got the highest average ADR in each year.

Conclusion: (Continue)

1. Its found that 77.3% of customers prefers BB(Bread & Breakfast).
2. Its observed that most of the customers rather guests do not repeat their bookings.
3. We can observe that in city hotels highest room ADR was for room type G and was lowest for the room type K. In case of Resort Hotels highest ADR was for room type H and was lowest for room type I.
4. It is observed that average daily count rate has decreasing trend after having peak value in August. This decreasing trend continues till month of January and after month of January ADR starts to increase and this trend is again observed till the month of August.
5. Each year ADR has been increasing consistently.